	Findings	Recommendations
	General Process Issues	
C	**Finding:** The committee is impressed and encouraged by EPA's progress, recognizing that the implementation of the recommendations in the NRC formaldehyde report is still in process. If current trajectories are maintained and objectives still to be implemented are successfully brought to fruition, the IRIS process will have become much more effective and efficient in achieving its basic goal of developing human-health assessments that can provide the scientific foundation for ensuring that risks posed to public health by chemicals are assessed and managed optimally.	**Recommendation:** EPA needs to complete the changes in the IRIS process that are in response to the recommendations in the NRC formaldehyde report and specifically complete documents, such as the draft handbook, that provide detailed guidance for developing IRIS assessments. When those changes and the detailed guidance, such as the draft handbook, have been completed, there should be an independent and comprehensive review that evaluates how well EPA has implemented all the new guidance. The present committee is completing its report while those revisions are still in progress.
C	**Finding:** Although it is clear that quality control (QC) of the IRIS assessment process is critical for the outcome of the program, the documents provided do not sufficiently discuss the QC processes or provide guidelines that adequately separate the technical methods from the activities of QC management and program oversight. For example, the role of the CASTs in the QC process is not specifically described.	**Recommendation:** EPA should provide a quality-management plan that includes clear methods for continuing assessments of the quality of the process. The roles of the various internal entities involved in the process, such as the CASTs, should be described. The assessments should be used to improve the overall process and the performance of EPA staff and contractors.
F		**Recommendation:** When extracting data for evidentiary tables, EPA should use at least two reviewers to assess each study independently for risk of bias. The reliability of the independent coding should be calculated; if there is good agreement, multiple reviewers might not be necessary.
F	**Finding:** The current scoping process for obtaining input from within the agency is clear, but opportunities for stakeholder input from outside EPA early in the process are less clear.	**Recommendation:** EPA should continue its efforts to develop clear and transparent processes that allow external stakeholder input early in the IRIS process. It should develop communication and outreach tools that are tailored to meet the needs of the various stakeholder groups. For example, EPA might enhance its engagement with the scientific community through interactions at professional-society meetings, advertised workshops, and seminars. In contrast, greater use of social media might help to improve communications with environmental advocacy groups and the public.

	Findings	Recommendations
F	**Finding:** EPA has taken steps to expand opportunities for stakeholder input and discussion that are likely to improve assessment quality. However, not all stakeholders with an interest in the IRIS process have the resources to provide timely comments.	**Recommendation:** Similar to other EPA technical-assistance programs, EPA should consider ways to provide technical assistance to under-resourced stakeholders to help them to develop and provide input to the IRIS program.
F	**Finding:** Promoting efficiency in the IRIS program is paramount given the constraint of inevitably shrinking resources. Thus, the committee agrees with EPA that stopping rules are needed given that the process for some IRIS assessments has become too long as revisions are repeatedly made to the assessments to accommodate new evidence and review comments.	**Recommendation:** The stopping rules should be explicit and transparent, should describe when and why the window for evidence inclusion should be expanded, and should be sufficiently flexible to accommodate truly pivotal studies. Such rules could be included in the preamble.
R		**Recommendation:** Regarding promotion of efficiencies, EPA should continue to expand its efforts to develop computer systems that facilitate storage and annotation of information relevant to the IRIS mission and to develop automated literature and screening procedures, sometimes referred to as text-mining.
F	**Finding:** The draft handbook and other materials are useful but lack explicit guidance as to the methods and nature of the use of expert judgment throughout the full scope of the assessment-development process, from literature searching and screening through integrating evidence to analyzing the dose-response relationship and deriving final toxicity values.	**Recommendation:** More details need to be provided on the recognition and applications of expert judgment throughout the assessment-development process, especially in the later stages of the process. The points at which expert judgment is applied should be identified, those applying the judgment should be listed, and consideration should be given to harmonizing the use of expert judgment at various points in the process.
	Problem Formulation and Protocol Development	
C	**Finding:** The materials provided to the committee by EPA describe the need for carefully constructed literature searches but do not provide sufficient distinction between an initial survey of the literature to identify putative adverse outcomes of interest and the comprehensive literature search that is conducted as part of a systematic review of an identified putative outcome.	**Recommendation:** EPA should establish a transparent process for initially identifying all putative adverse outcomes through a broad search of the literature. The agency should then develop a process that uses guided expert judgment to identify the specific adverse outcomes to be investigated, each of which would then be subjected to systematic review of human, animal, and in vitro or mechanistic data.

	Findings	Recommendations
F		**Recommendation:** For all literature searches, EPA should consult with an information specialist who is trained in conducting systematic reviews.
C	**Finding:** A protocol is an essential element of a systematic review. It makes the methods and the process of the review transparent, can provide the opportunity for peer review of the methods, and stands as a record of the review.	**Recommendation:** EPA should include protocols for all systematic reviews conducted for a specific IRIS assessment as appendixes to the assessment.
	Evidence Identification	
F	**Finding:** EPA has been responsive to recommendations in the NRC formaldehyde report regarding evidence identification and is well on the way to adopting a more rigorous approach to evidence identification that would meet standards for systematic reviews. This finding is based on a comparison of the draft EPA materials provided to the committee with IOM standards.	**Recommendation:** The trajectory of change needs to be maintained.
C	**Finding:** Current descriptions of search strategies appear inconsistently comprehensive, particularly regarding (a) the roles of trained information specialists; (b) the requirements for contractors; (c) the descriptions of search strategies for each database and source searched; (d) critical details concerning the search, such as the specific dates of each search and the specific publication dates included; and (e) the periodic need to consider modifying the databases and languages to be searched in updated and new reviews. The committee acknowledges that recent assessments other than the ones that it reviewed might already address some of the indicated concerns.	**Recommendation:** The current process can be enhanced with more explicit documentation of methods. Protocols for IRIS assessments should include a section on evidence identification that is written in collaboration with information specialists trained in systematic reviews and that includes a search strategy for each systematic-review question being addressed in the assessment. Specifically, the protocols should provide a line-by-line description of the search strategy, the date of the search, and publication dates searched and, as noted in Chapter 3, explicitly state the inclusion and exclusion criteria for studies.

	Findings	Recommendations
F		**Recommendation:** Evidence identification should involve a predetermined search of key sources, follow a search strategy based on empirical research, and be reported in a standardized way that allows replication by others. The search strategies and sources should be modified as needed on the basis of new evidence on best practices. Contractors who perform the evidence identification for the systematic review should adhere to the same standards and provide evidence of experience and expertise in the field.
R	**Finding:** One problem for systematic reviews in toxicology is identifying and retrieving toxicologic information outside the peer-reviewed public literature.	**Recommendation:** EPA should consider developing specific resources, such as registries, that could be used to identify and retrieve information about toxicology studies reported outside the literature accessible by electronic searching. In the medical field, clinical-trial registries and US legislation that has required studies to register in ClinicalTrials.gov have been an important step in ensuring that the total number of studies that are undertaken is known.
F	**Finding:** Replicability and quality control are critical in scientific undertakings, including data management. Although that general principle is evident in IRIS assessments that were reviewed, tasks appear to be assigned to a single information specialist or review author. There was no evidence of the information specialist's or reviewer's training or of review of work by others who have similar expertise. As discussed in Chapter 2, an evaluation of validity and reliability through inter-rater comparisons is important and helps to determine whether multiple reviewers are needed. This aspect is missing from the IOM standards.	**Recommendation:** EPA is encouraged to use at least two reviewers who work independently to screen and select studies, pending an evaluation of validity and reliability that might indicate that multiple reviewers are not warranted. It is important that the reviewers use standardized procedures and forms.
F	**Finding:** Another important aspect of quality control in systematic reviews is ensuring that information is not double-counted. Explicit recognition of and mechanisms for dealing with multiple publications that include overlapping data from the same study are important components of data management that are not yet evident in the draft handbook	**Recommendation:** EPA should engage information specialists trained in systematic reviews in the process of evidence identification, for example, by having an information specialist peer review the proposed evidence-identification strategy in the protocol for the systematic review.

	Findings	Recommendations
R	**Finding:** The committee did not find enough empirical evidence pertaining to the systematic-review process in toxicological studies to permit it to comment specifically on reporting biases and other methodologic issues, except by analogy to other, related fields of scientific inquiry. It is not clear, for example, whether a reporting bias is associated with the language of publication for toxicological studies and the other types of research publications that support IRIS assessments or whether any such bias (if it exists) might be restricted to specific countries or periods.	**Recommendation:** EPA should encourage and support research on reporting biases and other methodologic topics relevant to the systematic-review process in toxicology.
F	**Finding:** The draft preamble and handbook provide a good start for developing a systematic, quality-controlled process for identifying evidence for IRIS assessments.	**Recommendation:** EPA should continue to document and standardize its evidence-identification process by adopting (or adapting, where appropriate) the relevant IOM standards described in Table 4-1. It is anticipated that its efforts will further strengthen the overall consistency, reliability, and transparency of the evidence-identification process.
	Evidence Evaluation	
C	**Finding:** The checklist developed by EPA that is presented in the preamble and detailed in the draft handbook addresses many of the concerns raised by the NRC formaldehyde report. EPA has also developed broad guidance for the assessment of the quality of observational studies of exposed human populations and, to a smaller extent, animal toxicology studies. It has not developed criteria for the evaluation of mechanistic toxicology studies. Still lacking is a clear picture of the assessment tools that EPA will develop to assess risk of bias and of how existing assessment tools will be adapted.	**Recommendation:** To advance the development of tools for assessing risk of bias in different types of studies (human, animal, and mechanistic) used in IRIS assessments, EPA should explicitly identify factors, in addition to those discussed in this chapter, that can lead to bias in animal studies—such as control for litter effects, dosing, and methods for exposure assessment—so that these factors are consistently evaluated for experimental studies. Likewise, EPA should consider a tool for assessing risk of bias in in vitro studies.

	Findings	Recommendations
C	**Finding:** The development of standards for evaluating individual studies for risk of bias is most advanced in human clinical research. Even in that setting, the evidence base to support the standards is modest, and expert guidance varies. Furthermore, many of the individual criteria included in risk-of-bias assessment tools, particularly for animal studies and epidemiologic studies, have not been empirically tested to determine how the various sources of bias influence the results of individual studies. The validity and reliability of the tools have also not been tested. **Finding:** Thus, the committee acknowledges that incorporating risk-of-bias assessments into the IRIS process might take additional time; the ability to do so will vary with the complexity and extent of data on each chemical and with the resources available to EPA. However, the use of standard risk-of-bias criteria by trained coders has been shown to be efficient.	**Recommendation:** When considering any method for evaluating individual studies, EPA should select a method that is transparent, reproducible, and scientifically defensible. Whenever possible, there should be empirical evidence that the methodologic characteristics that are being assessed in the IRIS protocol have systematic effects on the direction or magnitude of the outcome. The methodologic characteristics that are known to be associated with a risk of bias should be included in the assessment tool. Additional quality-assessment items relevant to a particular systematic-review question could also be included in the EPA assessment tool.
R		**Recommendation:** EPA should carry out, support, or encourage research on the development and evaluation of empirically based instruments for assessing bias in human, animal, and mechanistic studies relevant to chemical-hazard identification. Specifically, there is a need to test existing animal-research assessment tools on other animal models of chemical exposures to ensure their relevance and generalizability to chemical-hazard identification. Furthermore, EPA might consider pooling data collected for IRIS assessment to determine whether, among various contexts, candidate risk-of-bias items are associated with overestimates or underestimates of effect.

	Findings	Recommendations
C		**Recommendation:** Although additional methodologic work might be needed to establish empirically supported criteria for animal or mechanistic studies, an IRIS assessment needs to include a transparent evaluation of the risk of bias of studies used by EPA as a primary source of data for the hazard assessment. EPA should specify the empirically based criteria it will use to assess risk of bias for each type of study design in each type of data stream. **Recommendation:** To maintain transparency, EPA should publish its risk-of-bias assessments as part of its IRIS assessments. It could add tables that describe the assessment of each risk-of-bias criterion for each study and provide a summary of the extent of the risk of bias in the descriptions of each study in the evidence tables.
F	**Finding:** The nomenclature of the various factors that are considered in evaluating risk of bias is variable and not well standardized among the scientific fields relevant to IRIS assessments. Such terminology has not been standardized for IRIS assessments.	**Recommendation:** EPA should develop terminology for potential sources of bias with definitions that can be applied during systematic reviews.
F	**Finding:** Although reviews of human clinical studies have shown that study funding sources and financial ties of investigators are associated with research outcomes that are favorable for the sponsors, less is known about the extent of funding bias in animal research.	**Recommendation:** Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessment.
F	**Finding:** An important weakness of all existing tools for assessing methodologic characteristics of published research is that assessment requires full reporting of the research methods. EPA might be hampered by differences in traditions of reporting risk of bias among fields in the scientific literature.	**Recommendation:** EPA should contact investigators to obtain missing information that is needed for the evaluation of risk of bias and other quality characteristics of included studies. The committee expects that, as happened in the clinical literature in which additional reporting standards for journals were implemented (Turner et al. 2012), the reporting of toxicologic research will eventually improve as risk-of- bias assessments are incorporated into the IRIS program. However, a coordinated approach by government agencies, researchers, publishers, and professional societies will be needed to improve the completeness and accuracy of the reporting of toxicology studies in the near future.

	Findings	Recommendations
C	**Finding:** EPA has not developed procedures that describe how the evidence evaluation for individual studies will be incorporated, either qualitatively or quantitatively, into an overall assessment.	**Recommendation:** The risk-of-bias assessment of individual studies should be carried forward and incorporated into the evaluation of evidence among data streams.
	Evidence Integration for Hazard Identification	
C	**Finding:** Critical considerations in evaluating a method for integrating a diverse body of evidence for hazard identification are whether the method can be made transparent, whether it can be feasibly implemented under the sorts of resource constraints evident in today's funding environment, and whether it is scientifically defensible.	**Recommendation:** EPA should continue to improve its evidence-integration process incrementally and enhance the transparency of its process. It should either maintain its current guided-expert-judgment process but make its application more transparent or adopt a structured (or GRADE-like) process for evaluating evidence and rating recommendations along the lines that NTP has taken. If EPA does move to a structured evidence-integration process, it should combine resources with NTP to leverage the intellectual resources and scientific experience in both organizations. The committee does not offer a preference but suggests that EPA consider which approach best fits its plans for the IRIS process.
C	**Finding:** Quantitative approaches to integrating evidence will be increasingly needed by and useful to EPA.	**Recommendation:** EPA should expand its ability to perform quantitative modeling of evidence integration; in particular, it should develop the capacity to do Bayesian modeling of chemical hazards. That technique could be helpful in modeling assumptions about the relevance of a variety of animal models to each other and to humans, in incorporating mechanistic knowledge to model the relevance of animal models to humans and the relevance of human data for similar but distinct chemicals, and in providing a general framework within which to update scientific knowledge rationally as new data become available. The committee emphasizes that the capacity for quantitative modeling should be developed in parallel with improvements in existing IRIS evidence-integration procedures and that IRIS assessments should not be delayed while this capacity is being developed.

	Findings	Recommendations
F	**Finding:** EPA has instituted procedures to improve transparency, but additional gains can be achieved in this arena. For example, the draft IRIS preamble provided to the committee states that "to make clear how much the epidemiologic evidence contributes to the overall weight of the evidence, the assessment may select a standard descriptor to characterize the epidemiologic evidence of association between exposure to the agent and occurrence of a health effect" (EPA 2013a, p. B-6). A set of descriptor statements was provided, but they were not used in the recent IRIS draft assessments of methanol and benzo[a]pyrene.	**Recommendation:** EPA should develop templates for structured narrative justifications of the evidence-integration process and conclusion. The premises and structure of the argument for or against a chemical's posing a hazard should be made as explicit as possible, should be connected explicitly to evidence tables produced in previous stages of the IRIS process, and should consider all lines of evidence (human, animal, and mechanistic) used to reach major conclusions.
F	**Finding:** EPA guidelines for evidence integration for cancer and noncancer end points are different; the cancer guidelines are more developed and more specific.	**Recommendation:** Guidelines for evidence integration for cancer and noncancer end points should be more uniform.
	Derivation of Toxicity Values	
C	**Finding:** EPA develops toxicity values for health effects for which there is "credible evidence of hazard" after chemical exposure and of an adverse outcome.	**Recommendation:** EPA should develop criteria for determining when evidence is sufficient to derive toxicity values. One approach would be to restrict formal dose-response assessments to when a standard descriptor characterizes the level of confidence as medium or high (as in the case of noncancer end points) or as "carcinogenic to humans" or "likely to be carcinogenic to humans" for carcinogenic compounds. Another approach, if EPA adopts probabilistic hazard classification, is to conduct formal dose-response assessments only when the posterior probability that a human hazard exists exceeds a predetermined threshold, such as 50% (more likely than not likely that the hazard exists).

	Findings	Recommendations
R	**Finding:** EPA has made a number of substantive changes in the IRIS program since the publication of the NRC formaldehyde report, including the derivation and graphical presentation of multiple dose-response values and a shift away from choosing a particular study as the "best" study for derivation of dose-response estimates.	**Recommendation:** EPA should continue its shift toward the use of multiple studies rather than single studies for dose-response assessment but with increased attention to risk of bias, study quality, and relevance in assessing human dose-response relationships. For that purpose, EPA will need to develop a clear set of criteria for judging the relative merits of individual mechanistic, animal, and epidemiologic studies for estimating human dose-response relationships.
R	**Finding:** Although subjective judgments (such as identifying which studies should be included and how they should be weighted) remain inherent in formal analyses, calculation of toxicity values needs to be prespecified, transparent, and reproducible once those judgments are made.	**Recommendation:** EPA should use formal methods for combining multiple studies and the derivation of IRIS toxicity values with an emphasis on a transparent and replicable process.
C	**Finding:** EPA could improve documentation and presentation of dose-response information.	**Recommendation:** EPA should clearly present two dose-response estimates: a central estimate (such as a maximum likelihood estimate or a posterior mean) and a lower-bound estimate for a POD from which a toxicity value is derived. The lower bound becomes an upper bound for a cancer slope factor but remains a lower bound for a reference value.
R	**Finding:** Advanced analytic methods, such as Bayesian methods, for integrating data for dose-response assessments and deriving toxicity estimates are underused by the IRIS program.	**Recommendation:** As the IRIS program evolves, EPA should develop and expand its use of Bayesian or other formal quantitative methods in data integration for dose-response assessment and derivation of toxicity values.
R	**Finding:** IRIS-specific guidelines for consistent, coherent, and transparent assessment and communication of uncertainty remain incompletely developed. The inconsistent treatment of uncertainties remains a source of confusion and causes difficulty in characterizing and communicating uncertainty.	**Recommendation:** Uncertainty analysis should be conducted systematically and coherently in IRIS assessments. To that end, EPA should develop IRIS-specific guidelines to frame uncertainty analysis and uncertainty communication. Moreover, uncertainty analysis should become an integral component of the IRIS process.
	C: relevant to current revisions of the process; F: related to future refinements of the methods; R: research related to the assessment process.	